Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 60 (2015) 1753 – 1762

Procedia Computer Science

19th International Conference on Knowledge Based and Intelligent Information and Engineering Systems

Fermi Estimate on the Web:

Placing Sensor Networks on the Web with Noise

Yoshiteru Ishida, Ryunosuke Chiba

Department of Computer Science and Engineering,
Toyohashi University of Technology
Tempaku, Toyohashi 441-8580, Japan

Abstract

This paper proposes a framework for web intelligence based on virtual sensor networks on the web. "Strong AI" requires a framework that allows the AI to make inferences using incomplete, dynamic and growing data on the web. An example of such inferences may be found in the human reasoning known as the "Fermi estimate". However, the huge amounts of data available through the internet necessitate an appropriate reasoning engine. The Fermi estimate is reformulated to be used in a virtual sensor network and several examples are discussed.

Peer-review under responsibility of KES International

Keywords: Fermi estimate; Frame problem; web mining; Internet of Things; open data; data assimilation; ontology; sensor networks; self-recognition model

1. Introduction

As the internet continues to grow, huge amounts of data are being archived in a scattered manner. It is important to be able to organize and integrate such data to enable them to be used for monitoring, analyzing and predicting global-scale events such as natural disasters and human disasters. The growth of data is accelerating for not only text data but also signal data from the Internet of Things (IoT) that connects sensors, actuators and control devices through the network. Thus, web data mining[5,6] needs to include data from the IoT.

To cope with this explosive growth of dynamic, large-scale and unformatted data, data sciences are becoming increasingly important. We have been studying a sensor network based on a self-recognition model that allows the sensor network to include not only measured data but also the credibility of data based on the relations among the measured data.[4]

On the other hand, to resurrect artificial intelligence (AI), AI must be strong enough to deal with incomplete and dynamic data, similar to flexible, robust human reasoning. The Fermi estimate (or Fermi problem) [1,11] is one such human reasoning named after the physicist Fermi, who was able to estimate such numbers as the number of piano tuners in an area using related available data but without knowing the actual data.

Nomenclature

$R_i(t)$	a credibility (normalized to a continuous value from 0 to 1) of a (virtual) sensor node i at time t.
T_{ij}	an evaluation from node i to j: 1 when node i evaluates node j as credible; −1 otherwise.

* Corresponding author. Tel.: +81-532-44-6895; fax: +81-532-44-6895.

1877-0509
Peer-review under responsibility of KES International
doi:10.1016/j.procs.2015.08.285

E-mail address: ishida@cs.tut.ac.jp

When we revisit the *Frame problem*[7] to make the reasoning mechanism stronger and more robust, we note that there could be two possibilities: the world model itself that the robot possesses is too weak, or even if the world model is strong enough the robot does not have a strong decision-making mechanism to choose adequate actions. We can identify something similar to the Frame problem in reasoning mechanisms on the web: the web itself does not reflect the real world, or even if it does, the reasoning mechanisms on the web are not strong enough to filter out "web noise". Here, we define web noise as redundant knowledge that can be filtered out by a reasoning mechanism such as those that can be fixed by weighting and mutual voting as opposed to non-weighting and simple flat voting (see Examples 4 and 6). Note that we do not expect the web to reflect the real world; we just want to know what the web is really saying.

In order to deal with the Frame problem on the web, the Fermi estimate should be reformulated for web data mining. For web data, which are not only expanding in scale but also encompassing qualitatively different data such as signals to and from sensors, actuators and control devices, this paper reports that a virtual sensor network based on the self-recognition model allows reasoning similar to the Fermi estimate on non-existing data or on data that include web noise, by building a virtual sensor network on the web. As an example of web noise, when we use a social media monitoring tool, we can get opposite results by subtly changing the input keywords. This reminds us of the Frame problem, even though it occurs for a software robot on the web.

Well-known frameworks that are used for web data processing include Page Rank[5], Bayesian networks[8], and Ontology networks.

Section 2 briefly introduces the Fermi estimate. Section 3 reviews the virtual sensor network on the web based on the self-recognition model. Section 4 reformulates the Fermi estimate as a framework for reasoning on web data, to fit the virtual sensor networks. Section 5 presents several examples of the virtual sensor network and the Fermi estimate on the network. The merits and demerits of the proposed framework are also discussed by comparing it with similar ones such as Bayesian networks.

2. Fermi Estimate and Sensor Networks

Fermi used an approximate calculation to estimate unknown numbers based on known and related numbers. On the other hand, the sensor network has been used for not only hard-wired sensor systems but also soft-oriented sensor systems including virtual sensors (composite sensors) that can be placed as web intelligent systems on the internet. Similar intelligent systems may be built on ontology-based reasoning on the web (semantic web or semantic reasoning). In order to apply the Fermi estimate to the soft-oriented sensor network on the web, we need to tailor the estimate. We characterize the Fermi estimate as:

- Target Data: The target numbers to be estimated are difficult to calculate or there are no exact data available on the web.
- Related Data: There are related numbers which are relatively easy to calculate or the data are available on the web.
- Relations: There are firm relations based on which the target numbers can be estimated from the related numbers.

Example 1. ("How many piano tuners in Chicago?"**)**[1]

Since the present study is inspired by the Fermi problem "How many piano tuners in Chicago?", we start with this example. The target data we want to know is the number of piano tuners in Chicago. As related data, we assume that we know the population of Chicago. A heuristic relation to obtain the target data from the related data is that the number of piano tuners is 1/30,000 of the population. The Fermi estimate is characterized as an approximate calculation of the heuristic relation with a chain of related data with several assumptions. Here, they are concatenated to a single relation to single related data (sensor node) for simplicity.

Other types of reasoning similar to human reasoning exemplified by the Fermi estimate can be found in higher order reasoning in qualitative reasoning such as order reasoning, dimension reasoning and approximate reasoning.

3. Virtual Sensor Networks based on Self-Recognition Model

3.1. Self-recognition model

This section reviews the self-recognition model on which a sensor network is constructed. The essential insight for using the self-recognition model for the virtual sensor network to integrate data on the web is that the data can be considered as agents which will have an active aspect and a certain autonomy (possibly their own logic of growth, i.e. selfishness). That is, the data on the web are not only to be evaluated or analyzed, but the data itself can play an active role in evaluating other related data, estimating non-existing data and interpolating missing data. We use the word “data integration” because these acts of evaluation, estimation and interpolation can be done mutually (two-directional, as opposed to one-directional), and even autonomously when a certain model is fitted to the web: the virtual sensor network with a virtual sensor node as the data on the web, and the relation among the nodes as the evaluation.

The self-recognition model (SRM, or self-recognizing network) consists of nodes capable of recognizing the states of other nodes: normal or abnormal. The results of recognition are indicated by the arcs from recognizing nodes to being-recognized nodes, and by the sign associated with the arcs: + when recognized as normal and – when abnormal. Recognition by abnormal nodes is unreliable.

These self-recognition models themselves can be mapped to a dynamic system called a dynamic relational network[4] or self-recognizing network[10] as a weighting and dynamic voting where the weight and vote change dynamically through feedback of the changing vote. Weighting the votes and propagating them identifies the abnormal nodes correctly under certain conditions. A continuous dynamic network is constructed by associating the time derivative of the state variable (expressing the vote) with the state variables of other nodes connected by the evaluation chain. The vote is normalized to a continuous value (called credibility) ranging from 0 to 1 to show the inferred results as a generalization of the binary value: 1 as true and 0 as false. Further, considering not only the effect from evaluating nodes but also that from evaluated nodes leads to the following dynamic system:

$$\frac{dr_i(t)}{dt} = \sum_j T_{ji}R_j + \sum_j T_{ij}R_j + \sum_{j\in\{k:\,T_{ik}\neq 0\}} (T_{ji}+1)$$

$$R_i(t) = \frac{1}{1+\exp(-r_i(t))}$$

where

R_i: credibility, which is the normalized value of r_i,

r_i: credibility before normalization,

T_{ij}: +1 (−1) for the arc from node i to node j with + (−) sign; 0 otherwise (for no arc).

In evaluating nodes, node j will stimulate (inhibit) node i when T_{ji} = 1 (−1). We call this model the black and white model, meaning that the network tries to separate an abnormal node clearly from a normal node; namely, the credibility (which differs from the probabilistic concept of reliability) of a node tends to be 1 (fully credible) or 0 (not credible), not an intermediate value. Moreover, we propose several different variants of this dynamic network such as the skeptical model and the gray model for different engineering needs. The results of this paper are generated only from the gray model, because we need the detailed quantitative information that results from the weighting and dynamic voting rather than binary results as explained in Example 2 below.

Although credibility looks like the mathematical concept of probability, the only shared aspect is that the value is normalized to 0 to 1. Credibility does not have mathematical rigor such as Bayesian networks[8]. For example, in mathematical probability (measure) all probabilities of all exclusive events covering the entire events must add up to 1, while credibility does not have even the concept of exclusive events. For computation of credibility, the only important point is consistency among credibilities of nodes. One way of computing credibility (or adjusting to the agent states) is to seek the credibility values that achieve the minimum consistency. This has been realized in the above black and white model by setting the total inconsistency analogous to energy (or a Lyapunov function) so that the total inconsistency decreases at each time step similarly to the Hopfield Network.

Example 2. (Liar’s Paradox)

As an illustration, let us explain the simplest example. In Figure 1 (*left*), if I say “I am a liar”, it causes a contradiction. If two people say “You are a liar” to each other (Fig. 1 *middle*), it is not possible to identify which one is actually a liar. However, if three people mutually evaluate each other as in the sign pattern shown in Fig. 1 (*right*), we can identify the liar.

This example illustrates that not only the node contents but also the relations (positive sign or negative sign) can include noise, and the noise can be filtered out by the credibility if the noise does not exceed a certain level.

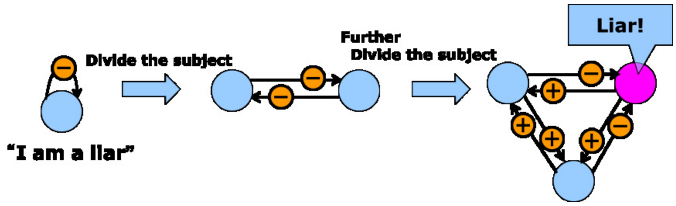

Fig. 1. Relaxing the liar's paradox by introducing mutual recognition in an autonomous distributed system. Nodes correspond to agents, and arcs with the + (−) sign mean the source agent says the target agent is honest (a liar).

The self-recognition model can be used for credibility reasoning not only for the credibility evaluation of each sensor node but also for the events that the sensors are monitoring. This credibility reasoning method can be compared with evidential reasoning based on the Dampster-Shafer theory[9] and probabilistic reasoning based on Bayesian networks.[8] As a reasoning method on web data, we compare it with Bayesian networks in Table 1. It should also be noted that the credibility of the self-recognition model can be computed in a fully autonomous and distributed manner at each node where the credibility of each node is computed by summing up the total weighted vote and propagating the computed credibility to the other related nodes.

Table 1. Comparison of Bayesian network and self-recognition model

Bayesian Network	Self-Recognition Model
Mathematical concept of probability related by the Bayesian rule	Credibility reflecting logical consistency among related sensor values
Requires initial probability to start propagation	Requires initial credibility to evaluate all the credibilities
Calculated in distributed and parallel manner	Calculated in distributed and autonomous manner
Dynamic data can be used in dynamic Bayesian networks	Dynamic data such as time series can be used (sensor values)
Mathematical rigor but complete data are required	No mathematical rigor but allows incomplete data

3.2. Virtual sensor networks and data assimilation

This section explores the sensor networks involving virtual sensors based on the self-recognition model. Let us call the virtual sensor networks involving virtual sensors as "data integration". As a data handling method, let us compare data integration (the virtual sensor networks, soft-oriented sensor networks or agent-oriented sensor networks on the web) with data assimilation (or model-based assimilation of observations[3]). However, the comparison is difficult because data assimilation encompasses several specific methods with a long history[1] whereas data integration covers more general reasoning methods. Starting from common parts, both data assimilation and data integration use consistency: the former on the models (mathematical models of dynamics) and the latter on the relation on the virtual sensor network. When space-time grids are used in the data assimilation and the virtual sensor node is placed in the space-time grids in the virtual sensor networks, both can be regarded as an interpolation and extrapolation of the space-time points, and further both can be used to detect anomalies not only in the observed data but also in the events to be observed.

One difference is that while data assimilation places importance on the mathematical model of dynamics, data integration focuses on the (approximate) reasoning based on the relations including heuristic ones. The approaches are briefly compared in Table 2. The following Example 3 may help to illustrate the reasoning component of data integration. Nevertheless, we admit that data assimilation can cope with the eight-coin puzzle and that data assimilation as a data handling method has been carefully tuned with mathematical rigor for large-scale numerical studies on space-time data such as in meteorology and oceanography.

Table 2. Comparison of data assimilation and data integration

Data Assimilation	Data Integration
Mathematical models of dynamics with space-time grids are used	Relations (including heuristics) with virtual sensor network are used
Numeric simulations on uniform space-time data can be done	Reasoning on qualitatively different data can be done
Large-scale space-time data may be assimilated to the model	Qualitatively different data may be integrated by the relation
Statistical model of Bayesian rule included	Comparable to Bayesian networks

As noted in the difference between data assimilation and data integration, the virtual sensor network can involve reasoning. One way to do this is dynamic activation of the relation. The following Example 3 of the virtual sensor network illustrates the reasoning parts.

Example 3. (Eight-Coin Puzzle)[4]

The eight-coin problem is a well-known puzzle: How many times must a balance be used to identify the one coin whose weight is heavier than the other seven coins? Let these coins be labeled 1, 2, 3, 4, 5, 6, 7, and 8 and grouped into {1, 2, 3}, {4, 5, 6}, {1, 4}, {2, 5}, {7}, {8}. The coin with the different weight can be identified by using a pair of balances only three times. However, here we focus on how the heavier coin can be identified by the sensor network.

Let $U(ij...k)$ denote the total weight of all the coins in the set $\{i, j, ..., k\}$. First, compare the weight between $U(123)$ and $U(456)$. If their weights differ, the mutual evaluation between the corresponding nodes is negative (not credible); and positive (credible) otherwise. This puzzle illustrates that a network can activate comparison depending on the credibilities of nodes.

$$\frac{dr_i(t)}{dt} = \sum_j \alpha_{ji} T_{ji} R_j + \sum_i \alpha_{ij} T_{ij} R_j - \frac{1}{2} \sum_i (\alpha_{ij} T_{ij} + 1)$$

where T_{ij} is the comparison result between $U(i)$ and $U(j)$: +1 for equal and −1 for not equal, thus $T_{ij} = T_{ji}$ for this example. The variable $\alpha_{ij}(t)$ (also assumed to be symmetrical in this example) is an activation level dependent upon the credibilities of nodes that activate or inactivate the relation from node i to j.

For simulation, the following are used:

$\alpha_{123,456}(t) = 1$,

$\alpha_{14,25}(t) = 1 - (R_{123}(t) + R_{456}(t))/2$,

$\alpha_{1,8}(t) = 1 - (R_7(t) + R_8(t))/2$

$\alpha_{1,7}(t) = 1 - (R_7(t) + R_8(t) + R_{14}(t) + R_{25}(t))/4$,

$\alpha_{2,7}(t) = 1 - (R_{14}(t) + R_{25}(t))/2$,

$\alpha_{3,7}(t) = (R_{14}(t) + R_{25}(t))/2$,

$\alpha_{7,8}(t) = (R_{123}(t) + R_{456}(t))/2$,

$\alpha_{6,7}(t) = (R_{14}(t) + R_{25}(t))/2$,

$\alpha_{4,7}(t) = 1 - (R_{14}(t) + R_{25}(t))/2$,

$\alpha_{5,7}(t) = 1 - (R_{14}(t) + R_{25}(t))/2$.

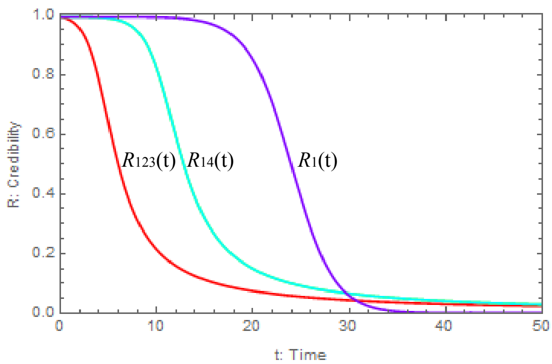

Fig. 2. Time evolution of credibilities corresponding to generalized nodes which monitor the value of the weight of the coin heavier than the others.

4. Web Intelligence by Virtual Sensor Networks on the Web

4.1. Reformulating the Fermi estimate for reasoning on the web and a reasoning algorithm

The Fermi estimate resembles higher order reasoning such as order based reasoning and qualitative reasoning of artificial intelligence. However, we focus on its management characteristic that even when the data involve noise or there is not enough data, the Fermi estimate manages to infer the non-existing data by finding existing data that have some relation with the target data. In order to apply this approach to the virtual sensor network, we reformulate the Fermi estimate as a strong reasoning (as in human reasoning) such that even without sufficient target data, it manages to find a sufficient amount of related data to estimate the target data using the relation between the target data and the existing data. For example, a popular problem for the Fermi estimate is: what is the approximate number of piano tuners in your state? Although you may not find the number on the web, if you can find on the web the number of pianos sold in the area and there is a strong correlation between the number of pianos sold and the number of piano tuners, then you can calculate the Fermi estimate of the number of piano tuners.

The Fermi estimate reformulated here thus looks like an extrapolation and interpolation of the target data, however, it can be used for prediction[10] too, for the future data is non-existing but may have a strong correlation with the current and past data.

The reasoning algorithm for the proposed Fermi estimate on the virtual sensor network is straightforward when the sensor network already exists: it evaluates the self-recognition model reflecting the existing data on the credibility of the nodes. This amounts to a dynamic weighting voting as opposed to flat voting of the hit numbers at each node. When the sensor networks do not exist for the target data, it should be built and maintained according to the following algorithm:

- If soft robots (crawlers) find the related data, then create new virtual nodes.
- Add the relations between the new nodes and existing nodes.
- If a node does not satisfy a certain condition, remove the node and the associated relations.

4.2. Primitives for deploying the virtual sensor network on the web

The (virtual) sensor network to be deployed on the web requires two primitives: the (virtual) sensor node and the relation between the nodes.

The sensor node should be simple enough so that each node will not include the web noise. Since the node is a sensor node, it senses some quantity. Since we are focusing on reasoning on the web, we restrict ourselves to the case where the sensor nodes directly monitor the quantity, including the hit numbers. Also, the credibility of the node is what will be computed and should reflect the question of interest. The following Example 4 clarifies the condition of the sensor node on the web.

Example 4. (Virtual sensor nodes on the web)

Suppose we do not know the grammatical rule regarding the indefinite article: "a weighting vote" or "an weighting vote". In this case, we can just ask some search engines to count the hit numbers, which turns out to be 7 to 0 by a search site of a domain in Japan on some date in 2015. Thus, the sensor network for this simple example turns out to be two nodes conflicting with each other. In this trivial case, the sensed value for the sensor node "a weighting vote" is 7 and that for the sensor node "an weighting vote" is 0. These sensed values are reflected in the initial credibilities of the sensor nodes. The relation between two sensor nodes is conflicting (mutually exclusive), and hence the sign of the arcs is minus in both directions. Opinions in opinion mining are defined to include the orientation expressing mutually exclusive words (or even neutral ones) with an orientation of positive, negative, or neutral.[6]

If the above example does not provide enough confidence, we can extend the network in time or space. That is, we let the virtual sensor nodes monitor for a longer time, for example, data acquisition once a month for a year. Perhaps a better way is to extend the network in space where the virtual sensor network is expanded using another type of relation, i.e. supporting relation.

Example 5. (Virtual sensor nodes related on the web)

In Example 4, the hit number 7 to 0 may seem too small and could include sample noise, so we could deploy two other sensor nodes: "a woman" and "an woman". Similarly to Example 4, the sensor nodes count the hit number at the present time: "a woman" as 299,000,000 and "an woman" as 326,000. It should be noted that both sensor nodes include some noise, for example, both sensor nodes may include the count of pages that explain why "an woman" should be "a woman". To extend the sensor network by combining the networks of Example 4, we introduce two mutually exclusive question nodes "a w… word" and "an w… word", and the supporting relation to these question nodes. The two pairs of nodes: "a weighting vote" and "a woman"; and "an weighting vote" and "an woman" are mutually supporting: hence the sign of the arcs is positive in both directions. Using the sensor network with these four sensor nodes added (Figure 3), it is concluded that "a w… word" wins the weighting and dynamic voting.

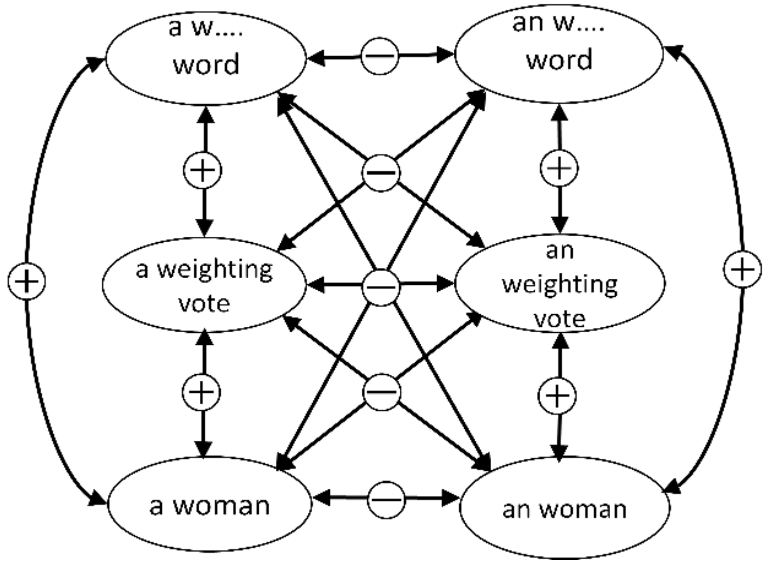

Fig. 3. A virtual sensor network for the question: *a* or *an* in front of a word starting with *w*?

However, we can realize that our true interest is in the higher knowledge as to whether '*a* in front of w starting a vowel sound' or '*an* in front of *w* starting a vowel sound'. Thus the virtual sensor nodes corresponding to these two are also deployed higher than the sensor nodes already deployed, which may be called instance node to the (higher) *question* nodes, and the relations between these two types of nodes are supporting from the instance nodes to the question nodes, that is, one directional with a positive sign. Figure 3 depicts the overall sensor network for this example. There are only two types of evaluation among nodes: mutually conflicting (exclusive), hence the sign of the arcs is minus in both directions; and supporting, hence the sign of the arcs is plus in both directions. These arcs with signs (evaluations) can be fixed when the sensor network is built, and remain fixed unlike the eight-coin puzzle (Example 3).

5. Fermi Estimate and Social Media Monitoring

5.1. Examples

In the following Example 6, an example of social media monitoring[6] (or social media listening) is illustrated by just two virtual sensor nodes which are mutually exclusive.

Example 6. (Social media monitoring)

Since we are interested in the current condition of the Japanese economy, we used a social media monitoring tool to check reputations appearing on the web on the date December 15, 2014. By using the input keyword KEIKI ("economy" in Japanese), we obtained relative hit counts (normalized from 0 to 1) including good evaluation 0.42 and bad evaluation 0.58, whereas when the input keyword KEIKI-DOKO ("economic trend" in Japanese) was used, we obtained the relative counts of 0.62 and 0.38.

The results are contradictory, so in the next Example 7 we deploy a virtual sensor network to filter out the possible web noise.

Example 7. (Fermi estimate under web noise)

In order to minimize web noise, we pose a more specific question than that of Example 6. We focus on the business outlook survey, for this is investigated and announced by the Japanese government in a specific period. Importantly, the survey is conducted by directly sending questionnaires to companies, which means we can check the results of the sensor network against the questionnaire survey. The question nodes are placed at the top (Fig. 4) in a mutually exclusive manner similarly to the sensor network (Fig. 3) of Example 5. The virtual sensor nodes are placed as three nodes which are mutually exclusive to each other and they are related to the top question nodes with a supporting link shown by a plus sign and a conflicting link shown by a minus sign, as shown in Figure 4. The measurements of these three virtual sensors are simple counts of the hit number using Google[5] in Japan (https://www.google.co.jp/) on some date in October 2014. The base hit number is counted with the base keywords: "business outlook survey months of 7 8 9 outlook for months 10 11 12". The sensor measurement of the node "Upward", for example, is the hit number with the base keywords and the word "Upward" added. The relative hit number (the hit number with the base keywords and "Upward") / (the hit number with the base keywords) is reflected as the initial credibility of the virtual sensor node "Upward". The initial credibility of other sensor nodes is similarly assigned except the question nodes whose initial credibility is set to 0.5.

The estimation results (credibility shown at each node in Figure 4) were: the economy is expanding 0.536, or shrinking 0.057 using the self-recognition model (gray model). The results are consistent with the Business Survey Index (BSI) of the duration: +0.5% point.

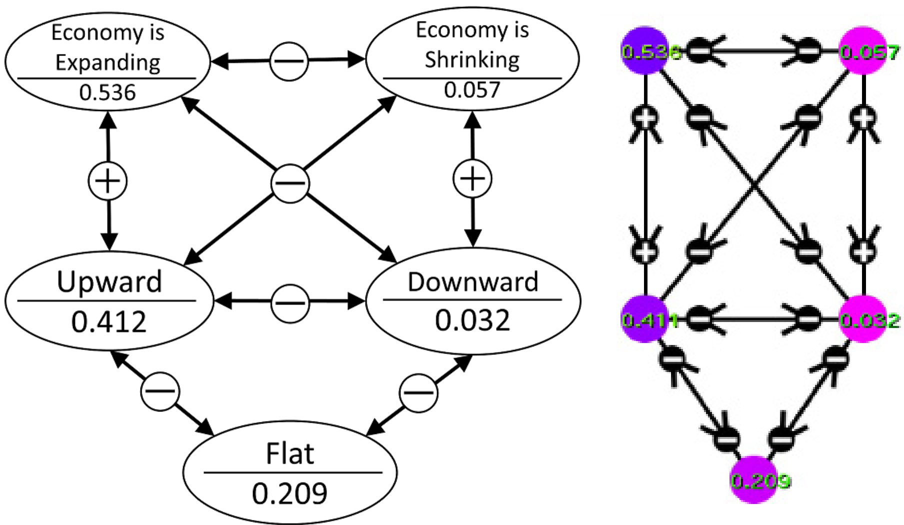

Fig. 4. A virtual sensor network for the question: Should only electric automobiles be used or not? Credibility of the top virtual node indicates that the Japanese economy is expanding 0.536, or shrinking 0.057. Illustrative figure (*left*) and snapshot of the sensor network (*right*).

We have tested several virtual sensor networks for questions such as: Should only electric automobiles be used or not in Japan? Is the recycling of Pet bottles good for the environment or not? The possible sensor network for the former question can be as complicated as shown in Figure 5. Although the reasoning on the network should reflect the web opinion more precisely than simple counting of the hit number, as is known from Examples 6 and 7, the reasoning results depend on the network structure and the way of measurement in each sensor node. The network structure and the sensor measurements require standardization as well as variations, for the appropriate network structure and the sensor measurement depend on the nature of the web data sensed.

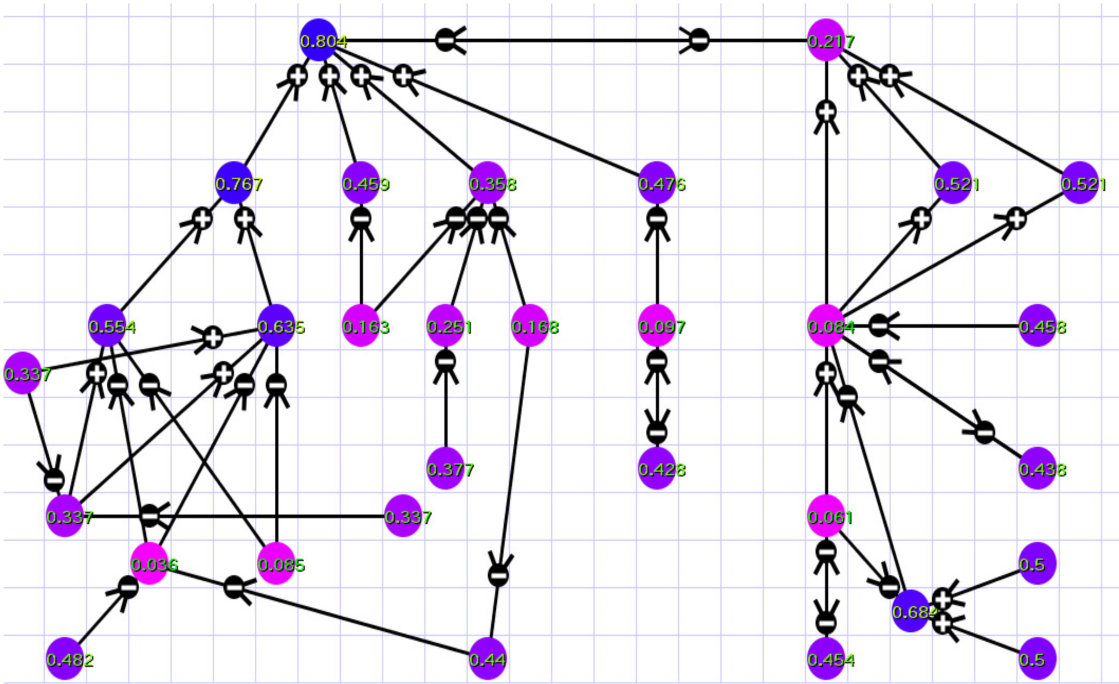

Fig. 5. A virtual sensor network for the question: Should only electric automobiles be used or not in Japan? Credibilities of the top virtual nodes indicate: affirmative 0.804, negative 0.217.

5.2. Types of sensor networks: Examples

Sensor networks involving virtual sensors and using the self-recognition model have been used in many areas, such as sensors and process diagnosis of chemical plants; monitoring of automobile combustion systems; diagnosis of motherboards, intrusion and abnormality detection for the smart home and space weather forecasting. Although most applications use data from hard-wired sensors, space weather forecasting uses remote sensing and *in situ* sensing data from satellites. However, if virtual sensor nodes could be placed on the internet, it would be possible to build a sensor network that can estimate unknown values of quantities of interest.

However, a sensor network and the Fermi estimate on the network only allow weighted and cyclic voting rather than the simple number of hits by search engines. Nevertheless, the possibilities are almost unlimited, and the system already has the ability to handle time series data (from IoT) and to visualize sensors or networked sensors with real-time sensed or estimated values together with credibilities calculated.

Let us show possible applications for several domains. For example, for the development of medical knowledge, huge amounts of medical data and life log data may be required. Although such data exist already and have been archived, the private nature of such data prevents usage without the consent of each person. If data exists with a strong correlation with the private data which do not have a privacy issue, and the Fermi estimate allows the private data to be estimated from the non-private data, then the sensor network could manage to work without directly using the private data.

Japan is aging rapidly and many local governments need to locate elderly people who wander about and become lost. Using the virtual sensor network, such people could be located not only by GPS-enabled smart phones but by estimating their location using an IoT-based sensor network through sensors placed in many facilities and shops. Similarly, a sensor network on the web could help victims after a natural disaster, prevent suicides and crime, and boost tourism by locating people and individuals. As is already done particularly in the electronic commerce market, huge amounts of data are extensively used to recommend products by estimating consumption behavior.

In short, the internet serves to exchange information similarly to the nervous system. Beyond this, the sensor network deployed on the web with IoT will serve to monitor events similarly to the immune system.[4]

5.3. Discussion

We have witnessed several large-scale growing systems that, once they start to grow, they develop their own logic of growth and become difficult to control; examples are money and the internet. Self-growing systems have both advantages and disadvantages. The upside is that they will grow by themselves and benefits of scale can be gained just by waiting for them to grow. However, the downside is that they will grow explosively, easily exceed the threshold and go out of control. Data, if given an active role, will grow and control will be lost. Mechanism design is a central issue of designing sound systems in the game theory of economics. We suppose that similar mechanism design may be needed before such systems become out of control.

Another point for data-centric science and engineering is its impact on explanatory science. The upside is that the results inferred or reasoned from huge amounts of data on the web have extraordinary powers of explanation and persuasion. The results can be used not only for scientific explanations but also for political ends or even advertising to gain high reviewer ratings, and can be used as the basis for decision-making at the individual or national level. The downside, of course, is the danger of arbitrariness of finding supporting data, as exemplified in Example 6 of social media monitoring. It is easy to find data that support your opinion even if the data are wrong or inappropriate. Many types of virtual sensor network should be studied to clarify what kind of sensor network is good for what kind of web noise.

6. Conclusions

We have shown that the Fermi estimate, when reformulated as human reasoning on incomplete data on the web, can be implemented on a virtual sensor network on the web. Future challenges are standardization and specification of the network structure and sensor measurement depending on the nature of the web data and the purpose of the estimate.

Acknowledgements

We did not cite the names of tools or specify the site search dates used in the examples, because this paper merely aims to illustrate how and why the virtual sensor network may be used on the web, rather than present specific values (retrieval hit numbers and credibilities) on the virtual sensor node. All keywords used for searching in the examples were Japanese; and were translated into English for illustrative purposes only.

References

1. von Baeyer HC. *The Fermi Solution*. Dover Publications; 1993.
2. Daley R. *Atmospheric Data Analysis*. Cambridge University Press; 1991.
3. Ide K, Courtier P, Ghil M, Lorenc AC. Unified Notation for Data Assimilation: Operational, Sequential and Variational. *Journal of the Meteorologcial Society of Japan* 1997; **75** 1B: 181–189.
4. Ishida Y. *Immunity-Based Systems: A Design Perspective*. Springer; 2004.
5. Langville AN, Meyer CD. *Google's PageRank and Beyond: The Science of Search Engine Rankings*. Princeton University Press; 2006.
6. Liu B. *Web Data Mining. Exploring Hyperlinks, Contents, and Usage Data*, Springer; 2007.
7. McCarthy J, Hayes PJ. Some philosophical problems from the standpoint of artificial intelligence. *Machine Intelligence* 1969*;* **4**: 463–502.
8. Pearl J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann; 1988.
9. Shafer G. *A Mathematical Theory of Evidence*, Princeton University Press; 1976.
10. Tokumitsu M, Ishida Y. A Space Weather Forecasting System with Multiple Satellites Based on a Self-Recognizing Network, *Sensors* 2014; **14** 5: 7974–7991.
11. Webb S. *If the Universe Is Teeming with Aliens... WHERE IS EVERYBODY*? Copernicus Books; 2002.